Mini-review: effect sizes and meta-analysis for antifouling research

Andrew J. Guerin* and Anthony S. Clare

School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

*Corresponding author. Email: andrew.guerin@newcastle.ac.uk

Mini-review: effect sizes and meta-analysis for antifouling research

Abstract

It is widely recognised that findings from experimental studies should be replicated before their conclusions are accepted as definitive. In many research areas, synthesis of results from multiple studies is carried out via systematic review and meta-analysis. Some fields are also moving away from Null Hypothesis Significance Testing, which uses *p* values to identify 'significant' effects, towards an estimation approach concerned with effect sizes and confidence intervals. This review argues that these techniques are underused in biofouling and antifouling research and discusses potential benefits of their adoption. They enable comparison of test surfaces even when these are not tested simultaneously, and allow results from repeated tests on the same surfaces to be combined. They also enable the use of published data to explore effects of different variables on the functioning of antifouling surfaces. Antifouling researchers should consider using these approaches and reporting results in ways that facilitate future research syntheses.

Keywords: research synthesis; systematic review; antifouling surfaces; surface properties; data analysis

Introduction

The term biofouling refers to the unwanted accumulation of biological organisms on surfaces. This can be a substantial problem in several contexts, most notably in dentistry and medicine (Harding & Reynolds 2014; Deshpande et al. 2015; Song et al. 2015), water handling and purification systems (Bogler et al. 2017; Jiang et al. 2017), and the maritime sector (Coutts & Taylor 2004; Gollasch 2002; Schultz 2007). Control of biofouling in these diverse contexts can be achieved using 'antifouling' surfaces, which use various mechanisms to reduce settlement, adhesion, and growth of biofouling organisms. The need for improvements to existing technologies means that antifouling research is a well-established and active sub-discipline; over 1000 relevant peer-reviewed articles were published in 2017 alone (Scopus database search for terms 'antifouling' OR 'anti-fouling', limited to 2017, conducted 15/06/2018). Many of these publications describe the design, formulation, and testing of antifouling surfaces. Typically, such tests can serve two functions: firstly they act as simple indicative tests of the antifouling characteristics of candidate surfaces, and secondly they can contribute to the fundamental understanding of biofouling and the factors underlying successful biofouling control.

With such a large body of literature, bringing together relevant research to create a coherent synthesis is challenging, particularly where studies are not in agreement. In common with other fields, knowledge is usually synthesised via 'narrative reviews'. However, narrative reviews suffer from a range of limitations and biases which have been discussed in the literature from other fields (Koricheva & Gurevitch 2013). Briefly, narrative reviews are not necessarily systematic in their survey of the literature, do not lead to quantitative estimates, usually do not have specified criteria for deciding which studies to include and which to exclude, and rely on personal judgement to arbitrate among papers with apparently contradictory findings.

In other fields this has led to the adoption of two complementary techniques: systematic review, and meta-analysis. Systematic review is the use of specific protocols for searching the literature, selecting studies, and harvesting the relevant data (Pullin & Stewart 2006); this means that the process can be repeated to verify findings and extend reviews with additional data. The term 'meta-analysis' refers more specifically to statistical methods for the analysis of data from multiple sources, enabling the integration of results from studies carried out at different times, even using potentially different methods. Modern meta-analysis methods mostly originated in the medical and sociological sciences (Hunt 1997), but are now becoming common in other fields, including ecological and environmental sciences (Stewart 2010). Underpinning meta-analysis methods are 'effect size measures', which describe the magnitude of the effect of a treatment relative to a control.

Null Hypothesis testing versus the 'estimation approach'

Conventional statistical analyses often follow the Null Hypothesis Significance Testing (NHST) approach. Briefly, this relies on the formulation of a statistical null hypothesis followed by a statistical test, which generates a *p* value. The difference between two samples is regarded as statistically significant if the calculated *p*-value falls below a preselected threshold level, α, which is usually set at 0.05.

Criticism of this approach is far from new (Rozeboom 1960; Cohen 1994) and there is a large and growing literature on the merits and limitations of NHST (Amrhein & Greenland 2018; Benjamin et al. 2018). This debate is beyond the scope of this review, but one relevant criticism of *p* values is that they are rather uninformative. They show whether or not any observed differences are 'statistically significant', but they provide no information about the size or importance of the observed differences (Gardner & Altman 1986). In the context of antifouling studies, they merely specify whether one surface performs better than another under a particular test regime. This does not necessarily aid detailed comparison of the relative performance of different antifouling surfaces, and *p* values cannot be related to surface properties in any meaningful way.

In contrast, the estimation approach is mainly concerned with estimating the magnitude of differences ('effect sizes') between different treatments, along with confidence intervals which represent the uncertainty associated with those estimates (Gardner & Altman 1986). This is sometimes referred to as 'the New Statistics' (Cumming 2012) even though the methods are not themselves particularly new (for example see Rothman 1978). These effect size measures and their confidence intervals are more informative than *p* values for comparing test results for different antifouling surfaces and understanding the effect of differences in surface properties.

Effect size measures for comparing test surfaces

Nearly any conceivable measure can be considered an 'effect size' in the broadest sense (Cumming 2012); for example, the mean proportion of barnacle cyprids settling on a surface, or the percent removal of biofilm on a surface exposed to shear stress. However, these types of measures are not useful for comparing antifouling properties except among surfaces that are tested simultaneously. This is because of several factors, including sampling error, variations in methodologies, differences in environmental conditions, differences in densities of cultured micro-organisms used in assays, and inherent variation in settlement behaviour. For example, barnacle cyprid settlement can vary substantially from one experiment to the next, even on identical surfaces (Figure 1).

[FIGURE 1 HERE]

To account for this variation among experiments, effect sizes are usually calculated based on the difference between a set of samples and some other 'control' group. Indeed, the term ‘effect size’ is usually used in this more restrictive sense. Fortuitously, many antifouling studies are conducted in a way that enables calculation of effect sizes, since it is common practice to include one or more standard surfaces for comparison. Most commonly, glass, polystyrene, or polydimethylsiloxane (PDMS) are included as standards. If samples tested on different occasions, or by different groups, are tested alongside the same standard surfaces, this can allow comparisons among studies. Although different types or batches of the ‘same’ standard surface may differ in some surface properties (different laboratories may use different PDMS blends, or create their standard surfaces differently), these variations could be accounted for. Side-by-side experimental comparison of different batches of a standard could allow simple adjustment of antifouling test data, or the potential use of slightly different standard types by different research groups could be accounted for in more sophisticated meta-analysis models (see below).

[TABLE 1 approximately here]

There are three main effect size measures that are likely to be useful for antifouling research (Table 1). The simplest is the ‘raw mean difference’: the difference between the means for a test surface and the chosen standard. This has the benefit that it is on the same scale as the original measurements, making it very simple to interpret. However, this measure can only be used when the range of possible values does not differ substantially among studies and where similar measurement methods and scales are used. In studies of diatom adhesion on surfaces, some authors use fluorescence intensity as a proxy for diatom density (eg. Sokolova et al. 2012), whereas others use direct counts of cell numbers (eg. Mieszkin et al. 2012); added to the usual expected variation among studies, this can lead to differences of as much as three orders of magnitude in diatom density measures. Clearly a simple comparison of mean differences is not appropriate in these circumstances.

The solution to this problem is the use of standardised mean differences. The most commonly used are the Cohen's *d* family of effect size measures, which are calculated by dividing the mean difference by the pooled standard deviation of the control and test samples. This expresses the difference between samples in multiples of the standard deviation and therefore represents a common currency for comparison among experiments even where measurement scales differ substantially among studies. The simplest version of *d* assumes that the standard deviations of the two groups are similar, but a version of *d* can be calculated in circumstances where the variances differ (Bonett 2008), or, more simply, the variance of either the sample or the standard surface can be used (Glass et al. 1981). *d* is often adjusted to account for small-sample bias; this adjusted value is usually referred to as Hedges' *g* (Hedges 1981; Cumming 2012). Unfortunately, naming conventions for effect size measures are somewhat variable (Cumming 2012; Lakens 2013), so it is always important to be clear about exactly what effect size measure has been used and how it has been calculated. For interpretation of *d* values, Cohen (1988) suggested benchmarks representing small (d = 0.2), medium (d = 0.5) and large (d = 0.8) effects. However, given the high effectiveness of some antifouling surfaces compared to standard surfaces (and therefore high values of *d*, see examples below), these default values are likely to be of little use in antifouling studies.

The other useful measure of effect size is the log mean ratio (also known as the response ratio); this is obtained by dividing the mean of the sample by the mean of the standard, and then applying a log transform (Hedges et al. 1999). Like *d*-family effect size measures, log mean ratios can be useful where different experiments use different measures or measurement scales.

There are many other possible effect size measures (Grissom & Kim 2005; Peng & Chen 2014), although most of these are less likely to be relevant for typical antifouling studies. Two very similar measures that are worth noting are the common language effect size

(McGraw & Wong 1992; Lakens 2013) and the probability of superiority (Grissom & Kim 2005). These give the probability that a randomly drawn replicate of a test surface performs better in a particular test than a randomly chosen replicate of the standard or control surface. They are easy to interpret, but are less useful for further analysis, since they reach an asymptote when all replicates of one surface out-perform all replicates of the other (when CLES or PS = 0 or 1), preventing further differentiation among better or worse performing surfaces. Formulas for calculating effect size measures and their confidence intervals are widely available in the literature (Nakagawa & Cuthill 2007; Cumming 2012; Kline 2013; Lakens 2013; Peng & Chen 2014; Cumming & Calin-Jageman 2016), but manual calculation (particularly of confidence intervals) can be complicated, so there are many software tools that can be used (Table 2) to facilitate this process.

[TABLE 2 APPROXIMATELY HERE]

If they have been calculated in the same way, effect sizes can be compared to assess relative performance of different test surfaces, formulations, or antifouling technologies. This may be particularly useful for 'down-selection' of coatings, where rapid testing of surfaces and selection for further tests is the priority. As an example, the SEAFRONT project (Synergistic Fouling Control Technologies, http://seafront-project.eu/) involved laboratory and field evaluation of a large number of coatings, followed by down-selection of a limited number of coatings for additional tests. Conversion of data from each antifouling assay into effect sizes allowed comparison of coating performance across the whole project. This informed the selection of a small number of candidate surfaces for further testing, but also provided additional insights. For assays involving the diatom *Navicula incerta* (Figure 2) many tested surfaces generated results that were similar to those for PDMS, but some technologies were clearly superior: one particular series of polymer hydrogels and a series of polymer films with incorporated zwitterions outperformed all other surfaces in this assay. However, the comparison also showed that the best performing surfaces were some of the earliest stage prototypes. All ten coatings that were significantly superior to PDMS (95% confidence intervals not crossing the zero line) were early stage prototypes, unsuitable for even short-term field testing. More mature, field-ready technologies did not show such good performance in this assay, illustrating the challenge of combining appropriate surface properties with the need to create robust, durable coatings for real-world application.

[FIGURE 2 HERE]

A small number of authors have already seen the value of using effect sizes (or similar ideas) to compare antifouling surfaces. Dobretsov & Thomason (2011) presented absolute and percent differences in the settlement of bacteria and diatoms on two test surfaces; Cordeiro et al. (2010) presented diatom adhesion density on two test surfaces as a percentage of that on a control surface; Stafslien et al. (2011) included 'normalised' differences in fouling-release properties relative to a standard surface across multiple assays; while Hibbs et al. (2015) explicitly used 'normalised' data to compare surfaces that were not tested simultaneously. All of these approaches demonstrate thinking in terms of effect sizes, and illustrate two points: firstly that effect sizes can be useful in antifouling research, and secondly that there is a lack of consistency in how they are applied. Each of these four studies used a different approach to normalisation, and in some cases it is not entirely clear how the normalised values (and the associated uncertainties, where present) were calculated.

The use of effect sizes is not limited to data from new experiments; they can be calculated using data from published studies. This is simple if the right information is clearly available: the mean, sample size and standard deviation for a test surface and an appropriate standard (control) surface. If some measure of variation other than standard deviation is supplied (standard error of the mean, 95% confidence intervals, etc), these can be converted into the standard deviation if the sample size and mean are reported. If one or more of these values are missing, effect sizes can still sometimes be calculated or inferred from *p*-values, *t*-values and other summary statistics (Lakens 2013). Some effect size measures (raw mean difference, ratio of means) can be calculated using only the means. However, without the variability and sample size, confidence intervals cannot be computed, which limits the value of the data and means that they cannot be used for full meta-analysis (see following sections).

Greatest speed and accuracy is facilitated if information is provided directly in tables or text, but visual methods can be used to extract means and error bar lengths from figures. Unfortunately, this can be extremely time-consuming, and requires that authors are clear about what their error bars represent; while identifying error bar types is regarded as good practice ("Rule 1" for use of error bars; Cumming et al. 2007) this basic information is often omitted.

Is inconsistent or unclear data presentation a problem in the antifouling literature? To investigate reporting practices, 100 relevant papers were obtained and examined to see whether they supplied sufficient information for effect size calculation (see Supplementary Material for details of search and selection protocol). Antifouling studies often include means, variability measures, and sample sizes for both test samples and an identifiable standard surface (Figure 3). However, only half of the studies examined directly provided sufficient information to calculate standardised mean differences. In most cases, information on means and variability measures was only provided graphically, or in a mixture of text and plots, with very few papers providing a complete set of tabulated data (Figure 3e,f).

[FIGURE 3 HERE]

Even so, a potentially large amount of data is available from the literature for comparison with new experimental data. There is, however, one important limitation to this approach. An effect size estimate based on one experiment still carries a high degree of uncertainty, since it is merely a single estimate of the true effect size. Given the low sample sizes often used in antifouling assays, this is an important consideration. It is perhaps less of a concern where the priority is rapid evaluation or screening of large numbers of prototypes, but it does mandate caution when interpreting data more broadly. It also underlines the importance of seeking, where possible, to replicate experiments. This is where meta-analysis becomes important.

Meta-analysis Part 1: integrating data from multiple tests

It is generally understood that the outcome of a single antifouling test of a surface is not definitive. The obvious solution is to repeat assays on multiple occasions when possible, but this leaves the challenge of how to interpret or combine results from repeat experiments, particularly if apparently contradictory outcomes are observed. Where assays have been repeated, it is fairly common practice for publications to simply report the results from a single 'representative' assay, rather than including all relevant experimental results. Meta-analysis provides a mechanism for combining the results of multiple tests to generate a single, combined effect size estimate, which is usually more precise than that obtained from single experiments. This also allows supposedly contradictory results to be reconciled as simply more or less accurate estimates of the true effect size. Several software tools exist which assist users in carrying out such simple meta-analyses (Table 2). These usually generate a 'forest plot' displaying the effect size measures and confidence intervals for every study or experiment, along with an estimated overall effect size with its own confidence intervals (Figure 4).

Aside from combining data from repeat experiments, meta-analysis can also be used to synthesise results from multiple published reports, papers and laboratory studies. To illustrate this, data from published papers and some additional laboratory data (see Supplementary material) were used to investigate the properties of three commercial fouling-release coatings: Intersleek® 700, Intersleek® 900, and Intersleek® 1100SR (produced by AkzoNobel), which are sometimes included in antifouling assays for comparison with test coatings. While it would be expected that these coatings will differ in real-life performance on ships (Intersleek® 1100SR is the newest generation and would be expected to be superior), it has been shown that *Navicula incerta* adheres strongly to low surface energy silicone elastomer coatings in laboratory assays (Holland et al. 2004). That this is also true for the Intersleek® 900 “fluoropolymer” and Intersleek® 1100SR “advanced fluoropolymer” coatings is borne out by the meta-analysis (Figure 4). For all three coatings, the overall estimated effect size is relatively close to zero, with confidence intervals clearly overlapping zero, indicating similar results to those of the standard (PDMS in this case). There also appears to be no substantial variation among the coatings. There is, however some variation among studies for all three surfaces. While much of the time the surfaces appear to perform similarly to PDMS, sometimes they perform better, and other times worse. While variation is expected as a result of sampling error, could known differences among experiments explain some of this variation and be accounted for in the meta-analysis?

In addition to combining results from multiple experiments, meta-analysis allows exploration of the influence of additional factors, referred to as moderators. For the percent removal data used in the above example, one possibility is that differences in the method of diatom removal could account for some of the variation in the study results. Removal experiments generally use one of two methods – exposure to shear stress (generally around 20-40 Pa) in a turbulent flow cell / flow channel (Schultz et al. 2000), or exposure to impact pressure from a water jet (Finlay et al. 2002), often at one of two pressures: 69 and 138 kPa (though other values are sometimes used). The type of removal test can therefore be included as a factor in the meta-analysis model (See Supplementary material), creating what is referred to as a mixed-effects meta-analysis model. For the demonstration data, the mixed-effects analysis reveals a significant effect of the moderator (experiment type) for Intersleek® 900 and Intersleek® 1100SR, but not for Intersleek® 700 (Table 3). Effect size estimates computed for each level of the factor (High pressure waterjet, Moderate pressure water jet, and Flow channel) show the relative ease of removal of diatoms from each coating in each type of test (Table 3, Figure 5). For all three coatings, when removal is tested using a flow channel or moderate impact pressure water jet, the surfaces are not distinguished from PDMS or each other. For the high pressure water jet, however, there appears to be a difference among the coatings. Intersleek® 1100SR performs significantly better than PDMS (CIs do not overlap zero) and appears to perform better than Intersleek® 700 and potentially Intersleek® 900.

Using the type of test as a moderator variable allowed separate estimates to be made for three different testing methodologies; an effect size estimate is thus available for the type of test that most closely matches any future experiment. Researchers carrying out similar tests of release of *N. incerta* from surfaces could use these estimates to compare the performance of their test surfaces against the expected release of *N. incerta* from these commercial fouling-release coatings.

This example has also illustrated one of the strengths of meta-analysis; the potential to gain additional insights that were not necessarily the aim of the analysis. Here, the goal was simply to integrate data from multiple experiments to generate better estimates of antifouling properties for some standard surfaces. Meta-analysis, however, also showed that some of the variability among experimental results came from differences in testing methods, and showed that the high pressure jet was able to resolve differences among surfaces that were not apparent using the other methods.

[FIGURE 4 HERE]

[FIGURE 5 HERE]

[TABLE 3 HERE]

Meta-analysis Part 2: meta-regression and effects of surface variables on antifouling

Meta-analysis, and in particular the use of moderators to explain variability among studies, can be taken further via meta-regression. Antifouling studies often include physico-chemical characterisation data, to verify that surfaces display the intended properties, and to explain differences in results among different coating formulations. Many factors are known to correlate with antifouling and fouling release performance, including (but not limited to): water contact angle and surface free energy (Finlay et al. 2010); surface polarity (Di Fino et al. 2014); elastic modulus (Sun et al. 2004); coating thickness (Wendt et al. 2006); surface roughness (Granhag et al. 2004); and presence and design of engineered surface structure (Schumacher et al. 2007). To date, most approaches to understanding the influence of these factors on biofouling have involved attempts to alter one variable while holding others constant. Such studies cannot entirely guarantee that no other surface properties, apart from the variable of interest, have been altered, and these can have substantial effects. For example, a recent study of polymer hydrogels (Ventura et al. 2017) appeared to show that the addition of sulfobetaine methacrylate, a zwitterionic material widely regarded as having good antifouling activity (Zhang et al. 2009; Aldred et al. 2010; Hibbs et al. 2015), reduced release of *N. incerta* under shear stress. Subsequent surface characterisation showed that this was most likely to be due to large changes in surface roughness, which were overwhelming the expected benefit of the presence of the zwitterion (Ventura et al. 2017).

Meta-analysis offers an additional means of exploring the effects of surface properties, since any measured physico-chemical variables are potential moderators for meta-regression models. This can allow use of much larger amounts of data, and comparison of the relative importance of multiple surface variables. To illustrate the potential use of meta-regression, an additional meta-analysis was conducted using data from published literature to explore the link between contact angle and adhesion of *N. incerta* (full methods in Supplementary material). Fourteen studies from the literature survey described above contained sufficient information on the percent removal of *N. incerta* to calculate effect sizes, along with contact angle data for tested surfaces. Meta-analysis identified a significant negative correlation between water contact angle and percent removal of *N. incerta* (Figure 6). This meta-analysis used unstandardised mean difference as the effect size measure; this is in the same scale as the original measurements so the relationship can be interpreted simply. The slope of the modelled relationship is -0.57 [95% CIs: -0.78, -0.36], indicating that the percent removal of diatoms from surfaces decreases by approximately 0.6 for every 1 degree increase in contact angle. This is in accordance with existing evidence that *N. incerta* adheres more strongly to hydrophobic surfaces. For example, Finlay et al. (2010) created a series of xerogels varying in contact angle, and tested adhesion of *N. incerta* to these surfaces. They observed the same trend – a negative correlation between percent removal and water contact angle. The slope of the relationship they identified was -0.65 (Figure 4a in Finlay et al. 2010), which is similar to that shown here, and well within the 95% confidence intervals of the meta-regression model.

[FIGURE 6 HERE]

Furthermore, the point at which the modelled relationship (Figure 6) crosses the x-axis is where the contact angle is 92° [95% CIs: 67, 119]. This is the contact angle value for which the model predicts performance to be the same as the standard surface (PDMS). Of the 14 studies used to provide data for this example analysis, six provided contact angle measurements for PDMS itself, giving a mean value of 103.8°; this is close to the model prediction.

This is a relatively simplified demonstration analysis, including only one moderator and ignoring other sources of variability, but it illustrates the usefulness of meta-analysis as a method of synthesising data from multiple studies to facilitate exploration of the effects of surface variables on antifouling. This approach could be applied to other surface variables, and other antifouling test data. The only limitation is the inconsistent reporting of relevant physico-chemical characterisation data. The same set of 100 studies used in the survey of data reporting (Figure 3) were used to examine trends in presentation of physico-chemical characterisation data in the biofouling literature (Figure 3g). Of the 100 studies examined, 83 provided some measures of physical or chemical surface properties (excluding data aimed only at verifying coating composition). Compared to the antifouling assay data, a much larger proportion (45 out of 83) reported surface characterisation data directly in text and tables rather than only in figures. Most studies provided contact angle data (80), while fewer than half of these used contact angles from multiple fluids to calculate surface energy (37). Smaller numbers reported any measure of surface roughness (30), elastic modulus (19) or thickness (14), while other surface properties were each reported in fewer than 10 papers. This means that it would be harder to gather sufficient data to examine relationships involving properties other than contact angle. It is of course entirely understandable that measurements are constrained by the availability and cost of relevant equipment, but researchers should be encouraged to try to characterise their surfaces as thoroughly as possible, to facilitate future meta-analysis.

Further considerations for meta-analysis

Systematic reviews and meta-analyses are only useful if they are well designed and carefully carried out, otherwise they risk suffering from the same problems as narrative reviews. Indeed, the existence of large numbers of poorly executed systematic reviews and meta-analyses is a problem in some disciplines (Ioannidis 2016). There are also assumptions underlying the statistical approaches used in a meta-analysis, and as with any statistical testing, the extent to which these assumptions hold should be considered. Many statistical methods assume independence of replicate data points, and similarly, standard meta-analysis assumes that each effect size measurement included in an analysis is an independent estimate of the 'true' effect size. If multiple effect size measurements from a single experiment, or single study, are included in a meta-analysis, then this assumption is technically violated. However, strategies exist for dealing with various types of statistical non-independence, the simplest being to either assume that the impact is negligible ('ignoring dependence'), or to only use one effect size per study ('avoiding dependence'); both of these options are obviously problematic (Van den Noortgate et al. 2013). Fortunately, there are also well-established modern statistical methods ('modelling dependence') which can be used with non-independent effect sizes (Gleser & Olkin 1994; Gleser & Olkin 2009; Konstantopoulos 2011; Van den Noortgate et al. 2013). To illustrate how some of these modelling techniques can be applied, they were used in the example meta-analyses presented above (see Supplementary material for full details). More recently 'cluster-robust variance estimation' methods have been developed, which can be used where there is statistical non-dependence, even when there is insufficient information to use some of the more complex approaches (Hedges et al. 2010; Fisher & Tipton 2015; Tipton & Pustejovsky 2015; Tanner-Smith et al. 2016). Importantly, these methods are somewhat easier to apply (Polanin et al. 2017), although they are only available for a limited range of statistical software.

A further issue that applies to meta-analysis of published material is the possibility of publication bias. Positive or exciting findings are more likely to be published, while negative, mundane, or corroborative findings are more likely to languish in the 'file drawer' and never be published (Joober et al. 2012). This presents a problem for meta-analysis based on published literature, since biased literature will potentially generate biased summary effect size estimates. However, since antifouling tests are often conducted simultaneously on a range of surfaces, including some which do not perform well, the field of antifouling research may be less prone to publication bias. In any case, there are methods to check for the presence of publication bias during meta-analysis (Egger et al. 1997; Sterne & Egger 2001).

It should be stressed that these concerns are more relevant to large meta-analyses and systematic reviews using data from multiple published studies, and should not deter researchers from carrying out their own small-scale meta-analyses to combine data from repeated experiments on the same surfaces. In any case, just as with primary experimental studies, meta-analyses should be comprehensively reported in order to facilitate repeatability and to justify any analytical decisions taken.

The zero problem

Particularly when testing high-performance antifouling surfaces, it is possible for means and variances of zero to be present in the data. For example, it is not unusual for studies of barnacle settlement to record zero settlement on at least one surface type (Berntsson et al. 2000; Dahlström et al. 2004; Fyrner et al. 2011). Zeros introduce two problems. Firstly, they create difficulties in the calculation of effect sizes. Log mean ratios cannot be calculated if one of the means is zero. If the mean for a sample is zero, standardised mean differences (eg. Cohens *d*) can be strongly inflated, particularly if the standard deviation is also low or zero. Aside from potentially creating exaggerated estimates of sample performance, this can lead to violation of meta-analysis model assumptions. The second problem with zero values is that a zero score in a settlement assay may not always have the same implications. For example, zero barnacle settlement in an experiment where settlement was generally low (eg. Figure 1, 9th Dec 2015) is not as impressive as zero settlement in an experiment where settlement was generally high (eg. Figure 1, 15th Sept 2015). A highly effective antifouling surface might have zero settlement at all times, but a less effective surface might have zero settlement on some occasions, and moderate settlement on others. Of course, this is an issue in all antifouling studies, and is not specific to meta-analysis or estimation approaches. Caution should always be exercised when interpreting zero values in antifouling data, and where possible, experiments should be repeated.

There are only limited solutions to this problem. If the standard deviation of the test sample is zero, then the standard deviation of the control surface could be used instead of the pooled standard deviation when calculating *d* (Glass et al. 1981). Alternatively, the raw mean difference could be used (if appropriate for the data). These options should prevent the occurrence of excessively inflated effect sizes, but do not address the underlying issues with zeros. It may appear tempting to simply exclude samples with zero values from meta-analyses, but this will reduce the amount of available data, and systematically excluding the best performing samples will introduce bias into any further analyses. For these reasons, such selective data removal is not appropriate. The final option is to conduct meta-analyses mainly using data from antifouling assays which are less likely to generate zero values. While zero settlement is relatively common in barnacle assays, tests using bacteria or diatoms usually record some adhesion of cells on surfaces.

Conclusions

The purpose of this article was to introduce antifouling researchers to the estimation approach and the use of meta-analysis methods. For those interested in pursuing these ideas, this review should be considered as a starting point, and the following sources are good next steps for those wishing to learn more: Nakagawa & Cuthill (2007); Borenstein et al. (2009); Cumming (2012); Kline (2013); Lakens (2013); Cumming & Calin-Jageman (2016). While this review has used marine antifouling examples to illustrate the ways in which effect sizes and meta-analysis could be useful, they are equally applicable to other biofouling contexts. There are a few challenges for the application of these approaches to biofouling data: the common presence of zeros values, the typically low sample sizes, and the inconsistent approach to physico-chemical characterisation of surfaces. On the other hand, there are some aspects that make antifouling data ideal material for meta-analysis. There is a large body of literature available, and reporting practises are generally good, although there is room for improvement in this regard. Importantly, calculation of effect sizes for comparison among studies is facilitated by the routine inclusion of standard surfaces. Finally, since antifouling tests are often conducted simultaneously on a range of surfaces, publication bias may be less of a problem than in other subject areas. It is still important for researchers to bear in mind that even 'disappointing' results from antifouling tests can be informative, as long as they accompany relevant surface characterisation data. Wherever possible, such data should be published to make them available to the wider research community.

Recommendations

(1) Antifouling researchers should consider reporting of effect sizes and confidence intervals, especially when these could be useful for comparing surfaces tested on separate occasions. This does not require any major changes to practice, particularly since these methods can be used alongside traditional presentation and statistical testing approaches.
(2) Where effect sizes and confidence intervals are reported, authors should always be clear about what effect size measures have been used and how they were calculated.
(3) Researchers should explore the use of simple meta-analyses as a tool for combining data from repeated tests on surfaces, allowing all assay data to contribute towards publications, rather than just reporting representative results.

(4) Meta-analysed effect sizes from multiple tests of standard surfaces could be used as general benchmarks for antifouling testing of experimental surfaces.

(5) Care should be taken over choices of effect size measures where data contain zeros, and consideration given to how best to analyse such data.

(6) Researchers should follow best practice by making sure that sample sizes are always reported, and that the type of error bars used in any plots is clearly stated. The ideal place for this information is in figure legends and data tables.

(7) All researchers should consider providing summarised data from all experiments (mean, sample size, standard deviation), or even raw data, as electronic supplementary material with their published papers, even if they could be inferred from figures in the paper. Given that this facility is widely available at no cost to authors, there is little reason to avoid this important contribution to 'open science' which would facilitate future meta-analytic work.

(8) Researchers should aim to broaden (as much as possible within equipment and funding limits) the range of physico-chemical characterisations carried out on samples, even if some may not seem directly relevant for specific studies.

(9) Whenever possible, data on 'poorly' performing surfaces should be published, since it can be just as informative as that on better performing surfaces.

Acknowledgements

The authors acknowledge financial support from the European Union Seventh Framework Programme via the SEAFRONT project under grant agreement number 614034. Drs M. Grainger, G. Stewart, K. Reynolds, A. Finnie, and two anonymous reviewers provided constructive comments which improved the manuscript. The authors also wish to thank all SEAFRONT project participants for the provision of prototype surfaces for testing, and for the many informative meetings and discussions that inspired this review.

Aldred N, Li G, Gao Y, Clare AS, Jiang S. 2010. Modulation of barnacle (*Balanus amphitrite* Darwin) cyprid settlement behavior by sulfobetaine and carboxybetaine methacrylate polymer coatings. Biofouling. 26:673–83. doi:10.1080/08927014.2010.506677.

Amrhein V, Greenland S. 2018. Remove, rather than redefine, statistical significance. Nat Hum Behav. 2:4. doi:10.1038/s41562-017-0224-0.

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, et al. 2018. Redefine statistical significance. Nat Hum Behav. 2:6–10. doi:10.1038/s41562-017-0189-z.

Berntsson KM, Andreasson H, Jonsson PR, Larsson L, Ring K, Petronis S, Gatenholm P. 2000. Reduction of barnacle recruitment on micro-textured surfaces: Analysis of effective topographic characteristics and evaluation of skin friction. Biofouling. 16:245–261. doi:10.1080/08927010009378449.

Bogler A, Lin S, Bar-Zeev E. 2017. Biofouling of membrane distillation, forward osmosis and pressure retarded osmosis: Principles, impacts and future directions. J Membrane Sci. 542:378–398. doi:10.1016/j.memsci.2017.08.001.

Bonett DG. 2008. Confidence intervals for standardized linear contrasts of means. Psychol Methods. 13:99–109. doi:10.1037/1082-989X.13.2.99.

Borenstein M, Hedges L V, Higgins JP, Rothstein HR. 2009. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons.

Cohen J. 1988. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge

Cohen J. 1994. The Earth Is Round ($p < .05$). Am Psychol. 49:997–1003.

Cordeiro AL, Pettit ME, Callow ME, Callow JA, Werner C. 2010. Controlling the adhesion of the diatom *Navicula perminuta* using poly(N-isopropylacrylamide-co-N-(1-phenylethyl) acrylamide) films. Biotechnol Lett. 32:489–95. doi:10.1007/s10529-009-0190-3.

Coutts ADM, Taylor MD. 2004. A preliminary investigation of biosecurity risks associated with biofouling on merchant vessels in New Zealand. New Zeal J Mar Fresh. 38:215–229. doi:10.1080/00288330.2004.9517232.

Cumming G. 2012. Understanding the new statistics. New York: Routledge.

Cumming G, Calin-Jageman RJ. 2016. Introduction to the New Statistics: Estimation, Open Science, and Beyond. New York, NY: Routledge.

Cumming G, Fidler F, Vaux DL. 2007. Error bars in experimental biology. J Cell Biol. 177:7–11. doi:10.1083/jcb.200611141.

Dahlström M, Jonsson H, Jonsson PR, Elwing H. 2004. Surface wettability as a determinant in the settlement of the barnacle *Balanus Improvisus* (Darwin). J Exp Mar Biol Ecol. 305:223–232. doi:10.1016/j.jembe.2003.12.013.

Deshpande A, Smith GWG, Smith AJ. 2015. Biofouling of surgical power tools during routine use. J Hosp Infect. 90:179–185. doi:10.1016/j.jhin.2015.03.006.

Dobretsov S, Thomason JC. 2011. The development of marine biofilms on two commercial non-biocidal coatings: a comparison between silicone and fluoropolymer technologies. Biofouling. 27:869–880. doi:10.1080/08927014.2011.607233.

Egger M, Smith GD, Schneider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical. BMJ. 315:629–634. doi:10.1136/bmj.316.7129.469.

Finlay JA, Bennett SM, Brewer LH, Sokolova A, Clay G, Gunari N, Meyer AE, Walker GC, Wendt DE, Callow ME, et al. 2010. Barnacle settlement and the adhesion of protein and diatom microfouling to xerogel films with varying surface energy and water wettability. Biofouling. 26:657–66. doi:10.1080/08927014.2010.506242.

Finlay JA, Callow ME, Schultz MP, Swain GW, Callow JA. 2002. Adhesion strength of settled spores of the green alga *Enteromorpha*. Biofouling. 18:251–256. doi:10.1080/08927010290029010.

Di Fino A, Petrone L, Aldred N, Ederth T, Liedberg B, Clare AS. 2014. Correlation between surface chemistry and settlement behaviour in barnacle cyprids (*Balanus improvisus*). Biofouling. 30:143–52. doi:10.1080/08927014.2013.852541.

Fisher Z, Tipton E. 2015. robumeta: An R-package for robust variance estimation in meta-analysis

Fyrner T, Lee H-H, Mangone A, Ekblad T, Pettitt ME, Callow ME, Callow JA, Conlan SL, Mutton R, Clare AS, et al. 2011. Saccharide-functionalized alkanethiols for fouling-resistant self-assembled monolayers: synthesis, monolayer properties, and antifouling behavior. Langmuir. 27:15034–15047. doi:10.1021/la202774e.

Gardner MJ, Altman DG. 1986. Confidence intervals rather than P values: estimation rather than hypothesis testing. Brit Med J. 292:746–750.

Glass GV, McGaw B, Smith ML. 1981. Meta-analysis in social research. Beverly Hill, CA: Sage Publications.

Gleser LJ, Olkin I. 1994. Stochastically dependent effect sizes. In: Cooper H, Hedges L V, editors. The Handbook of Research Synthesis. New York, NY: Russel Sage Foundation; p. 339–355.

Gleser LJ, Olkin I. 2009. Stochastically dependent effect sizes. In: Cooper H, Hedges L V, Valentine JC, editors. The Handbook of Research Synthesis and Meta-analysis. 2nd ed. New York, NY: Russel Sage Foundation; p. 357–376.

Gollasch S. 2002. The importance of ship hull fouling as a vector of species introductions into the North Sea. Biofouling. 18:105–121. doi:10.1080/08927010290011361.

Granhag L, Finlay J, Jonsson P, Callow J, Callow M. 2004. Roughness-dependent removal of settled spores of the green alga *Ulva* (syn. *Enteromorpha*) exposed to hydrodynamic forces from a water jet. Biofouling. 20:117–122. doi:10.1080/08927010410001715482.

Grissom RJ, Kim JJ. 2005. Effect sizes for research: a broad practical approach. Mahwah, NJ: Lawrence Erlbaum Associates.

Harding JL, Reynolds MM. 2014. Combating medical device fouling. Trends Biotechnol. 32:140–146. doi:10.1016/j.tibtech.2013.12.004.

Hedges LV. 1981. Distribution theory for Glass's estimator of effect size and related estimator. J Educ Stat. 6:107–128.

Hedges LV., Tipton E, Johnson MC. 2010. Robust variance estimation in meta-regression with dependent effect size estimates. Res Synth Methods. doi:10.1002/jrsm.5.

Hedges LV, Gurevitch J, Curtis PS. 1999. The meta-analysis of response ratios in experimental ecology. Source Ecol. 80:1150–1156.

Hibbs MR, Hernandez-Sanchez BA, Daniels J, Stafslien SJ. 2015. Polysulfone and polyacrylate-based zwitterionic coatings for the prevention and easy removal of marine biofouling. Biofouling. 31:613–624. doi:10.1080/08927014.2015.1081179.

Holland R, Dugdale TM, Wetherbee R, Brennan AB, Finlay JA, Callow JA, Callow ME. 2004. Adhesion and motility of fouling diatoms on a silicone elastomer. Biofouling. 20:323–329. doi:10.1080/08927010400029031.

Hunt M. 1997. How Science Takes Stock. The Story of Meta-analysis. New York, NY: Russel Sage Foundation.

Ioannidis JPA. 2016. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q. 94:485–514. doi:10.1111/1468-0009.12210.

Jiang S, Li Y, Ladewig BP. 2017. A review of reverse osmosis membrane fouling and control strategies. Sci Total Environ. 595:567–583. doi:10.1016/j.scitotenv.2017.03.235.

Joober R, Schmitz N, Annable L, Boksa P. 2012. Publication bias: What are the challenges and can they be overcome? J Psychiatr Neurosci. 37:149–152. doi:10.1503/jpn.120065.

Kline RB. 2013. Beyond Significance Testing: Statistics Reform in the Behavioral Sciences. 2nd Edition. Washington DC: American Psychological Association.

Konstantopoulos S. 2011. Fixed effects and variance components estimation in three-level meta-analysis. Res Synth Methods. doi:10.1002/jrsm.35.

Koricheva J, Gurevitch J. 2013. Place of Meta-analysis among other methods of research synthesis. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-analysis in Ecology and Evolution. Princeton, New Jersey: Princeton University Press; p. 3–13.

Lakens D. 2013. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Front Psychol. 4:1–12. doi:10.3389/fpsyg.2013.00863.

McGraw KO, Wong SP. 1992. A common language effect size statistic. Psychol Bull. 111:361–365. doi:10.1037/0033-2909.111.2.361.

Mieszkin S, Martin-Tanchereau P, Callow ME, Callow J a. 2012. Effect of bacterial biofilms formed on fouling-release coatings from natural seawater and *Cobetia marina*, on the adhesion of two marine algae. Biofouling. 28:953–68. doi:10.1080/08927014.2012.723696.

Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos. 82:591–605. doi:10.1111/j.1469-185X.2007.00027.x.

Van den Noortgate W, López-López JA, Marín-Martínez F, Sánchez-Meca J. 2013. Three-level meta-analysis of dependent effect sizes. Behav Res Methods. doi:10.3758/s13428-012-0261-6.

Peng C-YJ, Chen L-T. 2014. Beyond Cohen's *d*: alternative effect size measures for between-subject designs. J Exp Educ. 82:22–50. doi:10.1080/00220973.2012.745471.

Polanin JR, Hennessy EA, Tanner-Smith EE. 2017. A Review of Meta-Analysis Packages in R. J Educ Behav Stat. 42:206–242. doi:10.3102/1076998616674315.

Pullin AS, Stewart GB. 2006. Guidelines for systematic review in conservation and environmental management. Conserv Biol. 20:1647–56. doi:10.1111/j.1523-1739.2006.00485.x.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Rothman KJ. 1978. A show of confidence. New Engl J Med. 299:1362–1363.

Rozeboom WW. 1960. The fallacy of the null-hypothesis significance test. Psychol Bull. 57:416–428.

Schmid CH, Stewart GB, Rothstein HR, Lajeunesse MJ, Gurevitch J. 2013. Software for statistical meta-analysis. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of Meta-analysis in Ecology and Evolution. Princeton, New Jersey: Princeton University Press; p. 174–192.

Schultz MP. 2007. Effects of coating roughness and biofouling on ship resistance and powering. Biofouling. 23:331–41. doi:10.1080/08927010701461974.

Schultz MP, Finlay JA, Callow ME, Callow JA. 2000. A turbulent channel flow apparatus for the determination of the adhesion strength of microfouling organisms. Biofouling. 15:243–251. doi:10.1080/08927010009386315.

Schumacher JF, Carman ML, Estes TG, Feinberg AW, Wilson LH, Callow ME, Callow JA, Finlay JA, Brennan AB. 2007. Engineered antifouling microtopographies - effect of feature size, geometry, and roughness on settlement of zoospores of the green alga *Ulva*. Biofouling. 23:55–62. doi:10.1080/08927010601136957.

Sokolova A, Cilz N, Daniels J, Stafslien SJ, Brewer LH, Wendt DE, Bright F V, Detty MR. 2012. A comparison of the antifouling/foul-release characteristics of non-biocidal xerogel and commercial coatings toward micro- and macrofouling organisms. Biofouling. 28:511–23. doi:10.1080/08927014.2012.690197.

Song, Koo, Ren. 2015. Effects of material properties on bacterial adhesion and biofilm formation. J Dent Res. 94:1027–1034. doi:10.1177/0022034515587690.

Stafslien SJ, Bahr J, Daniels J, Christianson DA, Chisholm BJ. 2011. High-throughput screening of fouling-release properties: an overview. J Adhes Sci Technol. 25:2239–2253. doi:10.1163/016942411X574934.

Sterne JAC, Egger M. 2001. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. J Clin Epidemiol. 54:1046–1055. doi:10.1016/S0895-4356(01)00377-8.

Stewart G. 2010. Meta-analysis in applied ecology. Biol Lett. 6:78–81. doi:10.1098/rsbl.2009.0546.

Sun Y, Guo S, Walker GC, Kavanagh CJ, Swain GW. 2004. Surface elastic modulus of barnacle adhesive and release characteristics from silicone surfaces. Biofouling. 20:279–289. doi:10.1080/08927010400026383.

Tanner-Smith EE, Tipton E, Polanin JR. 2016. Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: a Tutorial in R. J Dev Life-Course Criminol. doi:10.1007/s40865-016-0026-5.

Tipton E, Pustejovsky JE. 2015. Small-Sample Adjustments for Tests of Moderators and Model Fit Using Robust Variance Estimation in Meta-Regression. J Educ Behav Stat. 40:604–634. doi:10.3102/1076998615606099.

Ventura C, Guerin AJ, El-Zubir O, Ruiz-Sanchez AJ, Dixon LI, Reynolds KJ, Dale ML, Ferguson J, Houlton A, Horrocks BR, et al. 2017. Marine antifouling performance of polymer coatings incorporating zwitterions. Biofouling. 33:892–903. doi:10.1080/08927014.2017.1383983.

Viechtbauer W. 2010. Conducting meta-Analyses in R with the metafor package. J Stat Softw. 36:1–48. doi:10.1103/PhysRevB.91.121108.

Wendt DE, Kowalke GL, Kim J, Singer IL. 2006. Factors that influence elastomeric coating performance: the effect of coating thickness on basal plate morphology, growth and critical removal stress of the barnacle *Balanus amphitrite*. Biofouling. 22:1–9. doi:10.1080/08927010500499563.

Zhang Z, Finlay JA, Wang L, Gao Y, Callow JA, Callow ME, Jiang S. 2009. Polysulfobetaine-grafted surfaces as environmentally benign ultralow fouling marine coatings. Langmuir. 25:13516–21. doi:10.1021/la901957k.

Table 1. Some effect size measures likely to be useful for antifouling studies. Position of sample versus standard in the equations is arbitrary and can be reversed as appropriate if the actual formula used is clearly communicated. For example, in a settlement-type assay, lower settlement indicates better antifouling performance, so the sample mean could be subtracted from the standard mean to give higher values for better antifouling performance.

Measure	Calculation	Advantages	Disadvantages
Raw mean difference	$\bar{x}_{sample} - \bar{x}_{standard}$	Simplicity Same scale as raw data	Inappropriate where mean differences vary substantially among experiments, or where different measures are used.
Standardised mean differences: 'Cohen's *d* family'	$\frac{(\bar{x}_{sample} - \bar{x}_{standard})}{pooled\ SD}$	Widely used. Useful where mean values differ widely among experiments, or where different measures are used.	Problematic when zero values present (or very low / zero variance) – eg. for very high performing surfaces.
Ratio of means (Response Ratio)	$ln\left(\frac{\bar{x}_{sample}}{\bar{x}_{standard}}\right)$	Simple to calculate Simple to interpret (after back-transform)	Not useful if either mean is zero.
'Common Language Effect Size' or 'Probability of superiority'	CLES: see Lakens (2013) PS: see Grissom and Kim (2005) or Peng and Chen (2014)	Simple to interpret and communicate.	Constrained between 0 and 1; no distinction among surfaces above a certain level of difference from the standard

Table 2. Some typical tools for calculating effect sizes and confidence intervals, and conducting meta-analysis. Hyperlinks active at time of publication. Values given for effect size measures (and particularly confidence intervals) may differ slightly for sources using slightly different algorithms. Researchers should use consistent and clearly identified means of effect size / CI calculation for data analysis. Some resources are intended as companions to a paper, book or course and are best used alongside these. Y = Yes, S = Sometimes, N = No.

Tool	Effect sizes	CIs	Meta-analysis		Advantages	Disadvantages	Examples
			Basic	Adv.			
Online 'calculators'	Y	S	N	N	Many available. Freely accessible. Simple to use.	May only allow single value to be calculated at a time. May not explain underlying formulae or calculation decisions. May not include confidence intervals on ES measures.	https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php
Downloadable 'calculators'	Y	S	N	N	Simple to use. Generally free. Some allow cut-and-paste data entry. Formulae can be copied.	Limited capabilities. Some may not allow cut-and-paste entry of large datasets	http://www.cem.org/effect-size-calculator https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00863/full - supplementary spreadsheet from Lakens (2013)
ESCI	Y	Y	Y	N	Good exploration of underlying concepts.	Comprehensive teaching tool best understood alongside accompanying books	Accompanies Cumming 2012, Cumming and Calin-Jageman 2016 https://thenewstatistics.com/itns/esci/
R statistical software (R Development Core Team 2013)	Y	Y	Y	Y	Highly flexible. Several meta-analysis packages available. Free and open-source. Comprehensive plotting capability. Advanced users can program additional functions.	Steep learning curve for new users	Example packages - ES calculation: *compute.es* Meta-analysis: *metafor* (Viechtbauer 2010) For a comprehensive list of meta-analysis packages see Polanin et al. (2017) or https://cran.r-project.org/web/views/MetaAnalysis.html
Other software	Y	Y	Y	S	Exact capabilities vary across multiple software solutions. Includes specialist meta-analysis software (eg. RevMan), and modules for general statistics programs (eg. SPSS). Cost varies widely (some are freely available). Dedicated user interfaces.		http://community.cochrane.org/tools/review-production-tools/revman-5 For detailed consideration of merits of different software packages for meta-analysis, see Schmid et al. (2013)

Table 3. Meta-analysis model outputs for Intersleek® data analysis, showing results for multivariate random effects models (no moderators) and multivariate mixed effects models with removal experiment type as a moderator (highjet = 138 kPa impact pressure; modjet = 69-81 kPa impact pressure; channel = 20-40 Pa shear stress). τ^2 estimates the total amount of heterogeneity in the model (for the random effects model) or the residual heterogeneity (for the mixed effects model). QM statistic is a test for statistical significance of the moderator (p < 0.05 indicates that the moderator is statistically significant). * these estimates have been included even though the moderator is not significant.

Surface	Random effects model		Mixed effects model including moderator (removal type)		
	τ^2	Estimate [95% CI]	τ^2	QM (df), *p*	Estimates [95% CI]
Intersleek® 700	4.87	-0.56 [-1.92, 0.81]	3.21	4.38 (2), *p* = 0.112	*highjet: 0.64 [-0.73, 2.01] *modjet: -2.01 [-4.31, 0.28] *channel: -1.12 [-3.29, 1.05]
Intersleek® 900	13.77	0.20 [-2.02, 2.43]	14.24	6.10 (2), *p* = 0.047	highjet: 2.87 [-0.22, 5.96] modjet: -0.52 [-4.16, 3.12] channel: -2.84 [-6.90, 1.22]
Intersleek® 1100SR	2.75	-0.082 [-1.16, 1.00]	1.64	26.04 (2), *p* < 0.0001	highjet: 5.53 [2.49, 8.57] modjet: 0.43 [-1.90, 2.75] channel: -0.52 [-1.21, 0.17]

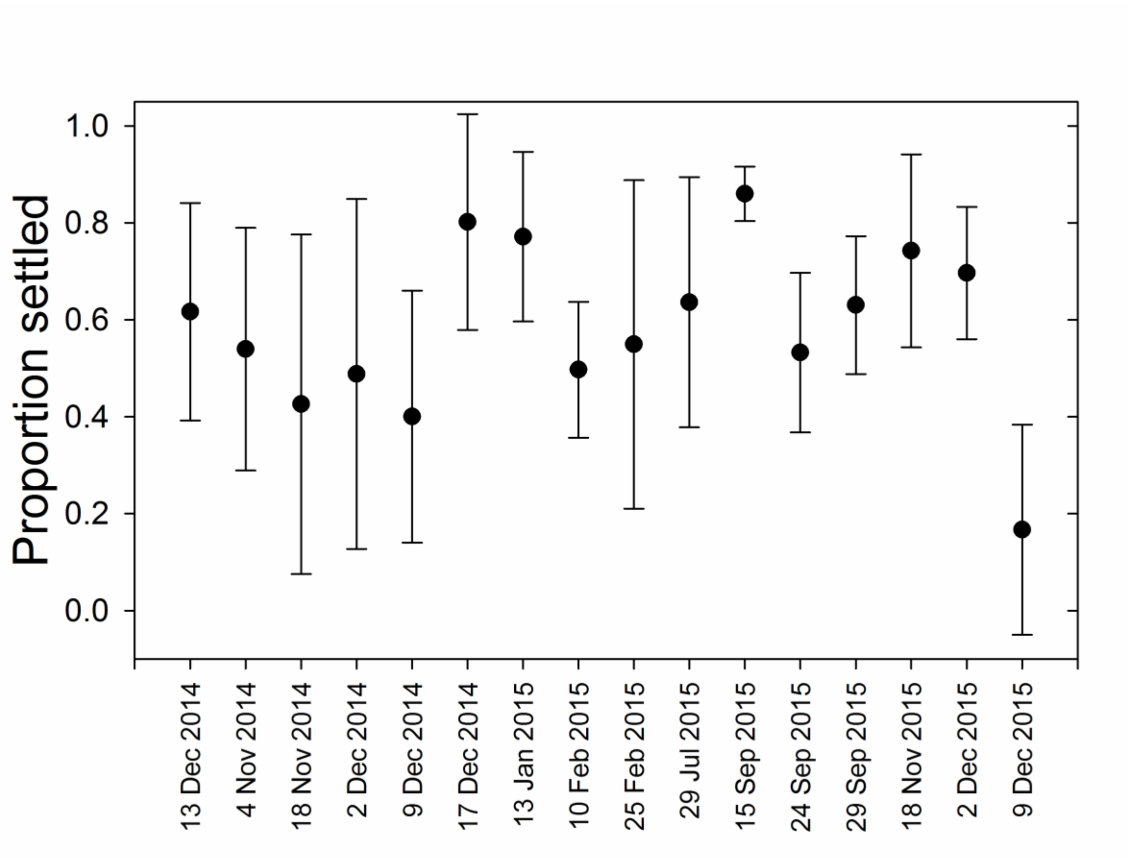

Figure 1. Variability in settlement of *Balanus amphitrite* cyprids on a consistent substrate, on 16 separate occasions. On each occasion, 20 (± 2) cyprids were pipetted in minimal 32 psu artificial seawater into each of 6 wells of a 24 well plate (costar® 24 well cell culture dish, Corning Inc. NY, USA), with each well containing 2 ml of 22 µm-filtered artificial seawater. Numbers of cyprids settled in each well were counted after 48 hours. Points show mean proportion settled in the 6 replicate wells, error bars are 95% confidence intervals.

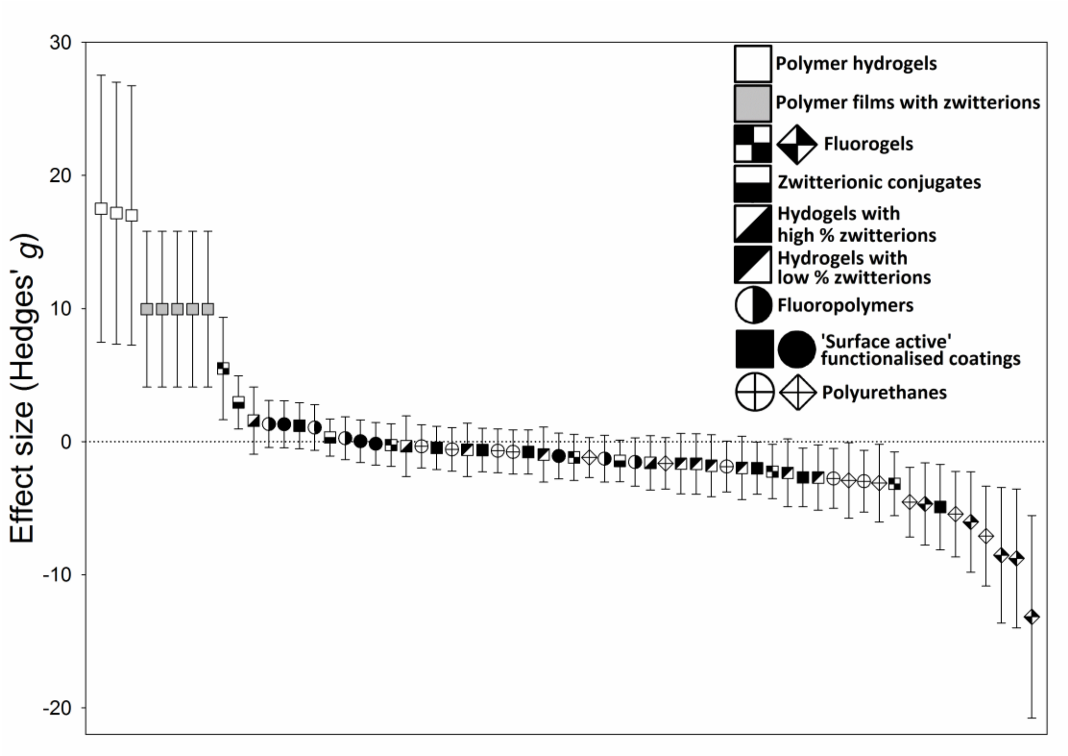

Figure 2. Effect size data for 62 candidate surfaces tested during the SEAFRONT project (see Supplementary material). Antifouling measure: diatom *(Navicula incerta)* density after exposure to 20-40 Pa shear stress in a flow cell (Schultz et al. 2000): these data were used to calculate effect sizes (Hedges' *g*) for each sample versus a standard surface (PDMS); error bars show 95% confidence intervals. Positive values indicate superior performance compared to PDMS (lower diatom density). Symbols indicate categories for comparison of different technology types; to protect intellectual property of SEAFRONT project partners, sample formulation details are not provided. Square symbols - early stage prototypes (unsuitable for field tests); Diamond symbols – improved / more robust prototypes suitable for field testing; Round symbols – advanced prototypes suitable for full field testing, and (hypothetically) real-world application. Dotted line at effect size zero represents equivalent performance to PDMS.

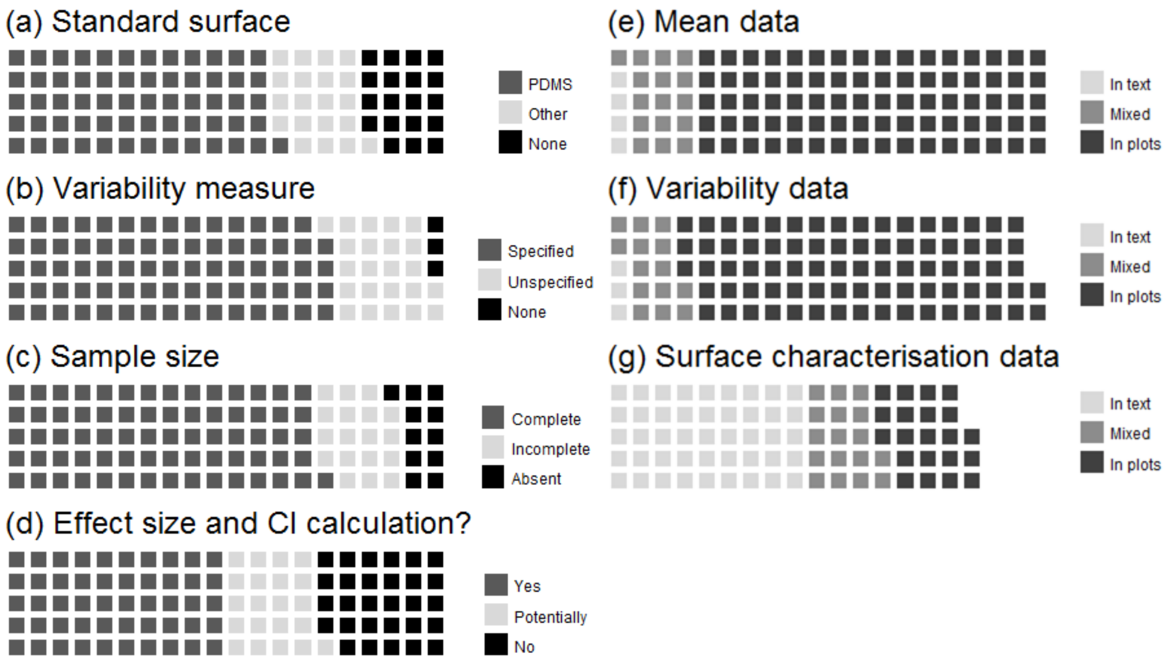

Figure 3. Data reporting behaviour from 100 antifouling studies, showing how many papers included (a) a standard surface, (b) clearly identified variability measures, and (c) sample size information. (d) shows how many papers presented enough information for effect size calculation: 'Yes' indicates a paper that provided enough information to calculate Hedges' *g* for at least one antifouling measure (mean, sample size and clearly identified variability measure for test surface and standard surface), 'Potentially' indicates that Hedges' *g* could have been calculated by making assumptions about the data (eg assuming a typical sample size where none was provided). Raw mean difference and mean ratios could have been calculated for most studies, but without sufficient information to generate confidence intervals. (e) (f) and (g) show how authors presented data on means, variability and surface characterisation (where present); "in text" refers to data provided in main text of published papers or in tables. Each block represents one published study.

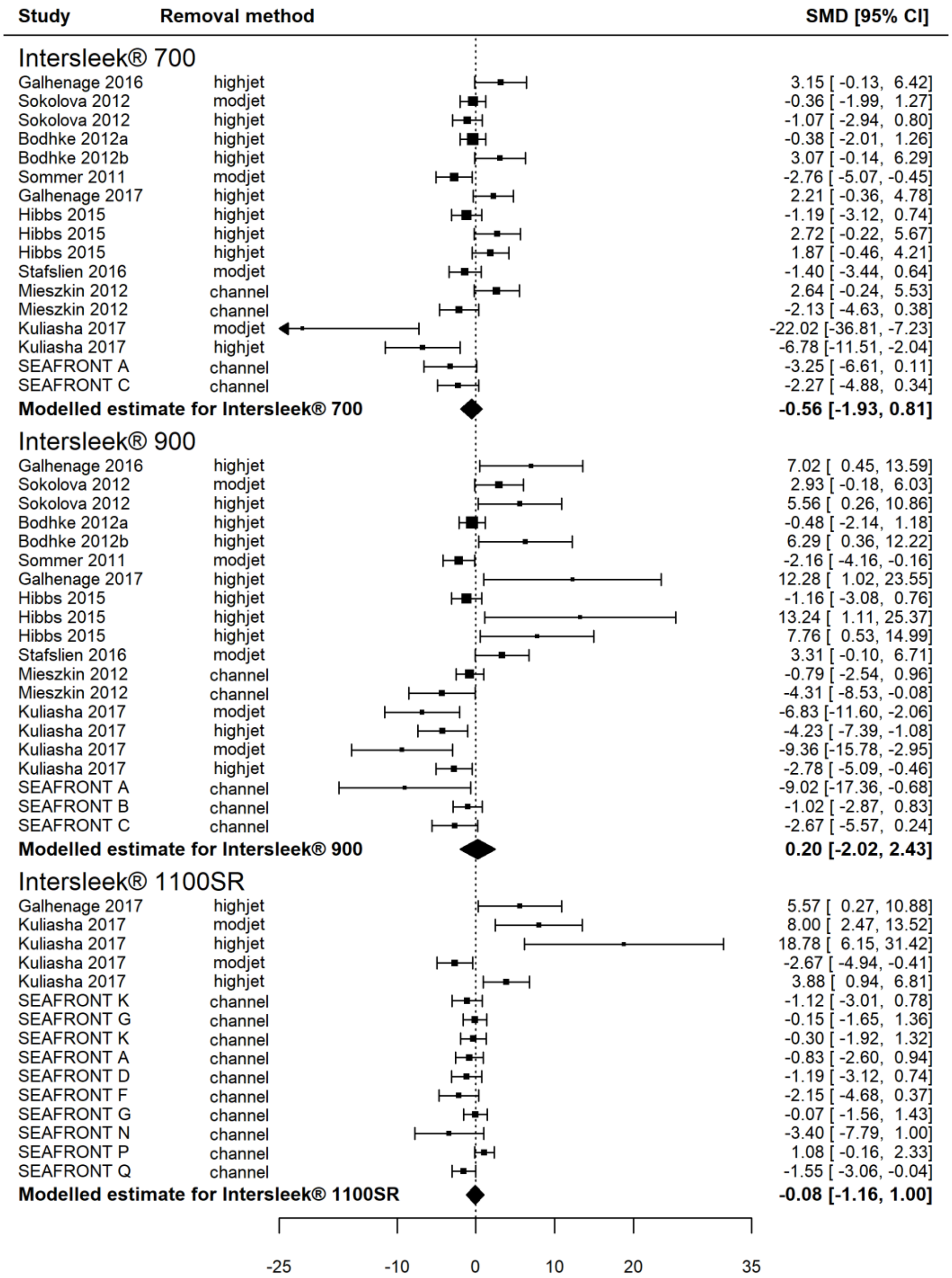

Figure 4. Forest plots showing meta-analyses of combined literature and SEAFRONT project data for antifouling performance of three commercial fouling release coatings, using PDMS as the standard surface. Underlying data are percent release of *N. incerta* diatoms during removal tests (see Supplementary material). Individual points show effect size estimates (SMD: Standarised Mean Difference, Hedges' *g*) and 95% confidence intervals for experiments from published studies. Effect size of zero (dotted vertical line) indicates equivalent performance to PDMS. Modelled effect size estimates for each coating were generated by multivariate random-effects model meta-analysis (details in Supplementary material), Different letters for SEAFRONT data denote identifiers for experiments conducted on different dates and have no other significance. Removal methods: highjet = 138 kPa impact pressure; modjet = 69-81 kPa impact pressure; channel = 20-40 Pa shear stress.

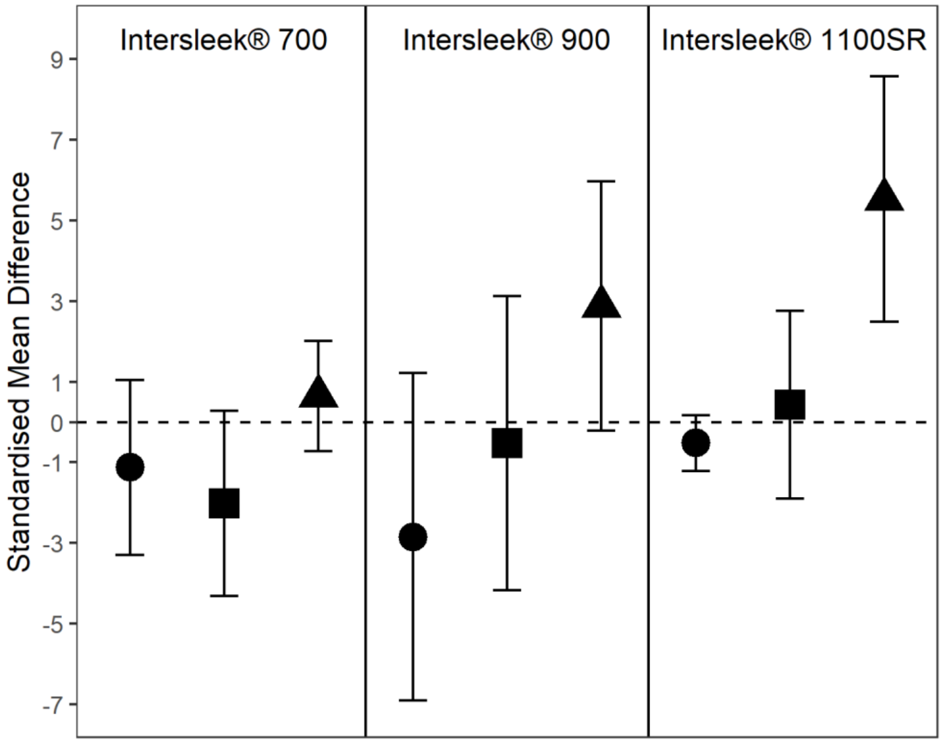

Figure 5. Combined effect size estimates (standardised mean difference, Hedges' *g*) generated by mixed-effects model multi-level meta-analysis (with test type as a moderator) for three testing methodologies: Triangle = High impact pressure ('highjet', 138 kPa); Square = Moderate impact pressure ('modjet', 69/81 kPa); and Circle = Shear stress in a turbulent flow channel ('channel', 20-40 Pa). An effect size of zero (dashed horizontal line) indicates equivalent performance to PDMS, error bars show 95% confidence intervals.

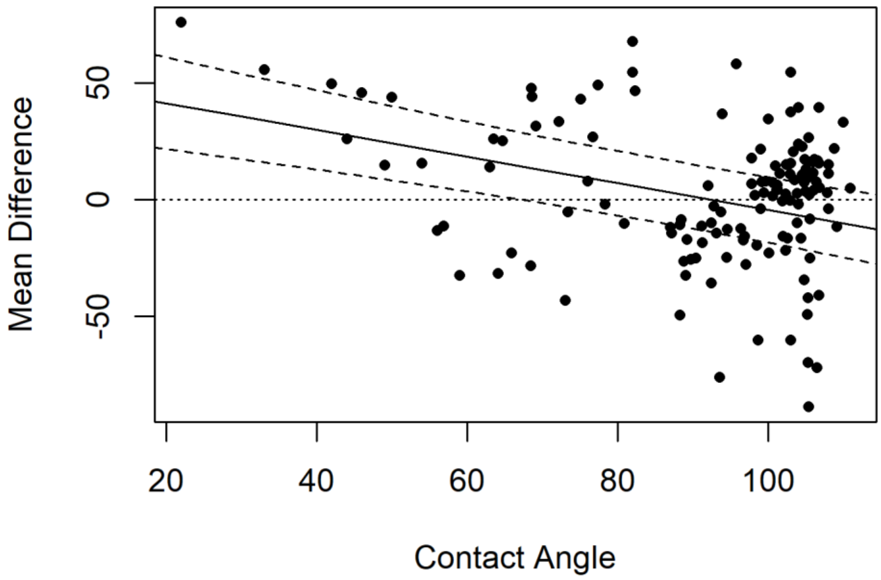

Figure 6. Meta-regression of contact angle and effect size (raw mean difference) for *Navicula incerta* percent removal data. Each data point is a surface tested in one of 14 studies (listed in Supplementary spreadsheet). The modelled relationship ($y = -0.5721x + 52.8615$) was significant, $p < 0.001$. Dashed lines are 95% confidence bounds for the modelled relationship.